

SCIENZA COGNITIVA E LIVELLI TEORICI

di Bruno Niceforo

Bruno Niceforo si è laureato in filosofia a Milano nel 1986 con Felice Mondella con una tesi sui più recenti orientamenti teorici e filosofici riguardo al problema mente-cervello. Docente di filosofia e storia nei licei, ha continuato ad interessarsi a questi problemi pubblicando alcuni articoli sui rapporti tra filosofia della mente e scienze cognitive. Si è poi occupato di comunicazione della scienza, specializzandosi in questo ambito presso la Scuola internazionale Superiore di Studi Avanzati di Trieste. Il presente articolo costituisce la sintesi di un lavoro prodotto durante il Corso di Perfezionamento in Epistemologia generale ed applicata organizzato dall'Università di Firenze. Niceforo.bruno@tiscali.it

INTRODUZIONE

In questo lavoro mi propongo di esaminare il problema dei rapporti tra diversi livelli teorici nel quadro della ricerca propria della Scienza cognitiva. Con questa espressione, come è noto, si fa riferimento ad un ambito multidisciplinare di ricerca, consolidatosi intorno alla metà degli anni settanta del secolo scorso, rivolto ad indagare la natura dei processi cognitivi avvalendosi dell'apporto di psicologia, neuroscienze, linguistica, intelligenza artificiale, filosofia della mente. L'elemento unificante di queste discipline è costituito dalla convinzione che i processi cognitivi possano essere spiegati attraverso modelli computazionali, cioè modelli che descrivono la cognizione in termini di elaborazione della informazione. E' anche noto che, a partire dalla metà degli anni ottanta, con l'emergere dei modelli connessionisti e poi delle ricerche sulla cognizione situata e sulla vita artificiale, si è creata all'interno di questo programma di ricerca una netta divaricazione di indirizzi: un primo indirizzo, talvolta denominato con l'espressione "scienza cognitiva classica", pur utilizzando anche i nuovi strumenti di ricerca, ha continuato a ritenere che la struttura portante dei processi mentali possa essere determinata da modelli che astraggono rispetto al substrato neurale che caratterizza gli organismi umani. Un secondo indirizzo (spesso denominato "scienza cognitiva post-classica"), insistendo con molta più determinazione sull'uso dei nuovi modelli, ha sostenuto che lo studio dei processi mentali non può in alcun modo astrarre rispetto al substrato neurale che li esprime e al contesto ambientale in cui essi si collocano. I due indirizzi divergono sia sul piano delle posizioni epistemologiche che su quello delle filosofie della mente che li ispirano.

In quel che segue, cercherò di delineare brevemente i fondamenti concettuali dei due suddetti indirizzi, per poi proporre un quadro di riferimento teorico (attualmente solo ad uno stato embrionale) distinto rispetto ai due appena citati, che mi appare più convincente e più ricco di potenziali sviluppi. Nella descrizione dei fondamenti teorici della scienza cognitiva classica farò riferimento soprattutto all'opera di Jerry Fodor; mentre, nella descrizione della scienza cognitiva post classica, mi appoggerò soprattutto alle tesi di Paul e Patricia Churchland. Gli autori in questione non esauriscono la gamma di posizioni riconducibili ai due indirizzi, ma costituiscono a mio parere la più chiara rappresentazione dei relativi punti di forza e di debolezza.

LA SCIENZA COGNITIVA CLASSICA

Il funzionalismo costituisce il retroterra filosofico del cognitivismo classico, consolidatosi a partire dagli anni sessanta (ma già anticipato negli anni cinquanta) in seguito alla crisi della tradizione comportamentista. Esso è caratterizzato innanzitutto da una prospettiva radicalmente antiriduzionistica riguardo al rapporto tra il livello teorico della psicologia e quello delle neuroscienze: il riduzionismo fisiologico, ovvero l'applicazione ai rapporti tra le suddette discipline del modello elaborato da Ernest Nagel (1960) viene criticato sulla base della distinzione tra fisicalismo delle occorrenze ("token physicalism") e "fisicalismo dei tipi" (type physicalism): secondo tale visione ogni evento mentale coincide con un evento cerebrale, ma l'identità si ferma ai singoli eventi e non può essere estesa alle proprietà mentali. L'argomento fondamentale contro il type physicalism, seguendo la efficace formulazione di Fodor (1975) prende le mosse dallo stato delle conoscenze in ambito neuroscientifico nei primi anni 70 nonché da alcuni riferimenti al pensiero biologico e approda poi alla *tesi della realizzabilità multipla dei processi mentali*.

I vari progetti di localizzare in modo preciso funzioni mentali nel cervello hanno incontrato molti ostacoli, data la plasticità esibita dall'organo cerebrale: quindi, dice Fodor, data una certa proprietà mentale, vi saranno molto probabilmente vari eventi neurologici ad essa correlabili. Né è plausibile, secondo Fodor, l'ipotesi che quegli eventi possano essere unificati sotto un unico predicato di una più matura teoria neuroscientifica, sul modello di quanto è avvenuto nei rapporti tra termodinamica e meccanica statistica: nel sostenere questo l'autore si appoggia ad alcuni lavori di filosofia della biologia (per esempio Hull, 1974) sui rapporti tra genetica mendeliana e biologia molecolare, che hanno mostrato non solo come fenomeni caratterizzati da

un singolo “predicato mendeliano” possano essere generati da diversi tipi di meccanismi molecolari, ma anche come uno stesso tipo di meccanismo molecolare possa generare fenomeni caratterizzabili solo nei termini di differenti predicati del livello “mendeliano”. Ora, la complessità dell’organo cerebrale non è certo inferiore a quella propria dei meccanismi biochimici interponentisi tra istruzioni del codice ereditario ed espressioni fenotipiche: risulta dunque assai probabile, secondo Fodor, che la molteplicità delle variabili coinvolte nel confronto tra predicati siti ai due livelli teorici (psicologico e neurologico) sia praticamente non controllabile, rendendo così impossibile l’identificazione.

Le precedenti osservazioni vengono poi rafforzate, secondo Fodor, dalla possibilità, ipotizzata dalle ricerche di Intelligenza Artificiale, che substrati fisici diversi dal cervello esplicino proprietà psicologiche analoghe a quelle degli organismi umani. E’ infatti ragionevole, secondo Fodor (anche alla luce delle ricerche di psicologia cognitiva, nonché delle ricerche linguistiche di matrice chomskiana) ammettere che i processi cognitivi propri della mente umana siano analizzabili nei termini di procedure effettive di calcolo. Ora, se si accetta la tesi di Church, risulta plausibile concepire la mente come un elaboratore che esegue computazioni secondo il modello di una macchina di Turing (o di qualche altro modello ad essa equivalente che, secondo Church, dovrebbe consentire di formalizzare la nozione intuitiva di calcolabilità). E poiché una macchina di Turing può avere realizzazioni fisiche del tutto diverse (essendo descrivibile in maniera completa al livello puramente logico delle sue istruzioni), altrettanto si potrà dire della mente umana. Ci sono dunque buone ragioni, secondo i funzionalisti, per pensare che la natura dei processi mentali umani sia indipendente da uno specifico supporto materiale e che sia possibile una descrizione completa degli stati mentali su un piano logico-informazionale. Questa è la tesi centrale del cognitivismo classico.

E’ evidente che questa prospettiva assume delle tesi non dimostrate (l’idea di analizzare i processi cognitivi in termini di procedure effettive di calcolo, per esempio, oppure la tesi di Church): tuttavia le congetture in questione hanno, secondo Fodor, una loro plausibilità dal punto di vista empirico, ed il programma di ricerca che esse implicano (la ricerca di modelli computazionali dei processi cognitivi) merita di essere perseguito.¹

¹ L’idea filosofica che la mente sia una macchina di Turing universale ha poi avuto una articolata traduzione, sul piano della teorizzazione psicologica, nei termini della *concezione modulare della mente*. La mente sarebbe insomma costituita da una rete assai complessa di dispositivi di analisi dell’informazione

La concezione epistemologica appena tratteggiata si salda poi, negli scritti di Fodor, con una teoria realistica dell'intenzionalità: cioè Fodor ritiene corretta l'idea, implicita nella psicologia del senso comune, che gli stati mentali si riferiscano ad oggetti e proprietà del mondo ed è convinto che anche una scienza cognitiva matura dovrà sviluppare generalizzazioni che quantifichino su stati interpretabili semanticamente. Si pone quindi il problema di come la dimensione semantica degli stati mentali sia conciliabile con l'idea centrale del funzionalismo: una macchina di Turing, infatti, opera esclusivamente sulla base di regole definite formalmente.

La strategia di Fodor per sciogliere questo nodo si basa su una riedizione della teoria rappresentazionale della mente di derivazione empiristica. *Fodor reinterpreta la nozione empiristica di rappresentazione concependola come un simbolo mentale all'interno di un linguaggio del pensiero* (detto anche "Mentalese"): esistono rappresentazioni atomiche (cioè stati funzionali espressi da una qualche forma di attività cerebrale) che si riferiscono a individui o proprietà del mondo; poi esistono rappresentazioni molecolari, ottenute composizionalmente, cioè combinando le rappresentazioni atomiche sulla base di proprietà formali che queste ultime esibiscono. La combinazione di queste rappresentazioni tra di loro può dare luogo a vere e proprie proposizioni mentali che hanno condizioni di verità (dipendenti dalla loro relazione con il mondo) e relazioni logiche con altre proposizioni.

Fodor formula poi la congettura che la teoria della dimostrazione sia in grado di formalizzare tutti i processi di pensiero realizzabili dalla mente umana; su questa base diventa possibile dare un fondamento rigoroso all'idea che la manipolazione compiuta dalla macchina cerebrale sui simboli del Mentalese, pur rispondendo esclusivamente alle loro proprietà formali, possa conservarne le proprietà semantiche. La semantica dei simboli mentali risulterebbe così implicita nella loro sintassi e quest'ultima è perfettamente "istanziabile" in un sistema fisico, consentendo così di inscrivere i fenomeni intenzionali in un quadro naturalistico senza ridurli a proprietà del cervello.²

altamente specializzati (i moduli, appunto, ognuno assimilabile ad una specifica macchina di Turing): tralascio qui le differenze nella specifica articolazione di questa teoria (per Fodor, per esempio, i moduli sono dispositivi *incapsulati informativamente*, e la concezione modularistica non è applicabile alla sfera del pensiero; mentre altri psicologi la pensano diversamente). Il punto importante, è che *l'architettura modulare della mente è pienamente individuabile al livello funzionale-computazionale che abbiamo tracciato fin qui.*

² Naturalmente resta il problema di come i simboli primitivi del linguaggio del pensiero possano godere di proprietà semantiche. Come è noto, la proposta di Fodor (1987) va nella direzione di quella che è stata chiamata una *semantica causal-informativa*, soluzione ingegnosa, che comporta però diversi problemi (vedi per es. Casalegno 1998).

LA SCIENZA COGNITIVA POST-CLASSICA

A partire dalla metà degli anni '80 la ricerca nell'ambito della Scienza cognitiva è stata attraversata da profondi mutamenti, che hanno portato alcuni autori alla convinzione che fosse in gioco un mutamento di paradigma in senso kuhniano. Le cause di questo mutamento sono legate a diversi fattori:

- 1) la ricerca nell'ambito delle reti neurali e sulla vita artificiale (vedi Churchland P. 1995; Churchland P.S., Sejnowski T.J.1992)
- 2) la ricerca relativa alla "cognizione situata" (vedi Brooks R. 1991, Van Gelder T. 1999).

La ricerca sulle reti neurali è tornata al centro dell'interesse di molti scienziati cognitivi (dopo la pionieristica opera di Rosenblatt) a metà degli anni ottanta, (vedi Rumelhart e McClelland 1986). L' applicazione di questi modelli allo studio dei processi cognitivi comporta, almeno in prima battuta, una netta contrapposizione rispetto all'approccio classico fin qui delineato: le computazioni effettuate dalla rete, infatti sono assai più vicine alle caratteristiche dei sistemi nervosi esistenti in natura rispetto a quelle proprie di un computer seriale: l'informazione su cui la rete opera non è rappresentata in maniera esplicita attraverso unità discrete, *ma è distribuita lungo tutta la rete*, e viene elaborata non secondo le procedure seriali di una macchina di Turing, ma in parallelo. Come è noto, oggi è possibile modellizzare attraverso reti o congiunzioni di reti diversi meccanismi psicologici: dal riconoscimento di facce alla lettura di un testo, al costituirsi di una rappresentazione di tipo concettuale o proposizionale ai processi di elaborazione che presiedono alla memoria episodica o semantica.

Molti ricercatori, per la verità, concepiscono questa metodologia di ricerca come una strategia complementare all'approccio funzionalista sopra delineato: le reti neurali avrebbero lo scopo di chiarire il funzionamento di quei moduli i cui compiti sono talmente specifici da non poter più essere sottoposti ad analisi funzionale: in altre parole, l'analisi funzionale del comportamento umano, (di per sé svincolata, nel suo apparato concettuale, dalla necessità di aderire al substrato neurale) avrebbe lo scopo di delineare la *macrostruttura* del mentale ed ai modelli connessionisti resterebbe il compito di individuarne la *microstruttura*.

Ma vi è un'altra linea interpretativa, per la quale la ricerca sulle reti neurali viene inserita in un quadro più ampio, dove un ruolo determinante è anche svolto dalle ricerche sulla nuova robotica e sulla vita artificiale.

La prima, in particolare, ha sottolineato il *carattere situato della cognizione*, progettando “agenti adattativi autonomi” (come i “Mobot” di R. Brooks). Il cosiddetto “frame problem”, ovvero il problema di dotare un agente artificiale di tutta l’informazione necessaria ad affrontare le situazioni reali in un formato che sia gestibile in tempo reale, viene affrontato con una sorta di rivoluzione concettuale rispetto alla visione “rappresentazionalista” della Intelligenza Artificiale classica: quest’ultima astrae completamente dalla presenza del corpo, concepito come una sorta di meccanismo esecutore di ordini progettati e codificati dai vari algoritmi mentali.³ Ma l’intelligenza umana, secondo l’approccio che stiamo descrivendo, non opera mediante uno schermo rappresentazionale definibile magari, come nel caso di Fodor, come insieme di proposizioni in “Mentalese” che si riferiscono allo stato delle cose del mondo; c’è un *irriducibile elemento di dipendenza contestuale* nei nostri comportamenti che può essere spiegato solo attraverso *l’attività anticipatoria di un corpo che è costantemente situato nell’ambiente* e che usa gli oggetti stessi come modelli per la sua attività.

I “mobot”, anziché costruire complessi modelli dell’ambiente, agiscono direttamente in esso selezionando quel limitato ambito di informazioni che sono stati progettati per trattare. La loro flessibilità risulta essere molto maggiore ed il loro modo di operare, pur richiedendo capacità discriminative e di azione non superiori a quelle di un insetto, sembra molto più vicino dei modelli tradizionali ai comportamenti locomotori tipicamente umani. Detto in altri termini, la conoscenza umana non è interpretabile in termini di “know that”, ma nei termini di “know how”. E’ questa una linea che ha ispirato anche le ricerche di J.J. Gibson, che ha costantemente sottolineato il carattere *diretto* della percezione umana, cioè il fatto che gli organismi umani (e non solo) non utilizzano modelli dettagliati del mondo, ma sono predisposti a rilevare “affordances”, ossia possibilità di uso, intervento ed azione che l’ambiente locale offre.

Su questa base teorica viene posto in discussione il cuore dell’approccio classico, cioè l’ipotesi che esista un livello “simbolico” di organizzazione dei processi mentali (il cosiddetto “inconscio cognitivo”) distinto dal piano dei processi neuronali e

³ Non a caso Fodor presentava il *solipsismo metodico* come la più corretta cornice filosofica in cui inserire le proprie teorie: la mente è un motore sintattico che opera nella più completa ignoranza del significato dei simboli su cui opera (pur conservandolo, come abbiamo visto); ogni nostro processo mentale avviene come se gli altri uomini e la natura intorno a noi potessero essere messi tra parentesi. Una prima critica a questa visione (sulla linea appena esposta) era stata formulata da Hubert Dreyfus (1972); con la grossa differenza che Dreyfus ha difeso l’opportunità di salvaguardare il *piano fenomenologico* di descrizione dei vari eventi mentali (per nulla riducibile, secondo lui, a quello dei processi neuronali), mentre la Scienza cognitiva post-classica ha scelto una prospettiva eliminativistica.

descrivibile in maniera indipendente dalle variabili ambientali. Nel caso dei più radicali sostenitori (vedi Churchland P. 1995) si giunge poi ad una prospettiva *eliminativistica*: si ritiene cioè che il piano “fenomenologico” di descrizione degli eventi mentali (ed i fenomeni collegati alla coscienza e all’intenzionalità che in esso si esprimono) sia destinato a dissolversi in un complesso sistema di reti neurali integrate tra di loro, situate nell’ambiente nel modo sopra descritto ed evolutesi nel corso del tempo secondo linee definite da algoritmi genetici. Credenze e desideri verranno sostituite da vettori di attivazione su insiemi di unità connessioniste⁴

Si potrebbe obiettare che molti nostri comportamenti sono causati dall’attribuzione ad altri di determinate credenze (le cosiddette credenze di secondo ordine: come quando ci comportiamo in un certo modo verso qualcuno perché crediamo che quella persona la pensi in un certo modo): la folk psychology, pur nella inevitabile approssimazione del suo apparato concettuale, sembra assai difficile da eliminare senza una notevole perdita di potere esplicativo. Si risponde riducendo credenze e desideri a modalità (imprecise) con cui quei sistemi di reti neurali evolute che sono gli organismi umani hanno cominciato a descrivere i propri processi interni. Esse (credenze e desideri) costituiscono ancora oggi lo strumento con cui interpretiamo le azioni nostre e degli altri; ma il fatto che esse abbiano un’utile funzione strumentale nei processi sociali non implica affatto che esse costituiscano un modo scientificamente attendibile per spiegare i nostri comportamenti.

POSSIBILI LINEE DI INTEGRAZIONE TRA SCIENZA COGNITIVA CLASSICA E POST-CLASSICA.

Gli autori che presento in questa ultima parte non condividono né la prospettiva fodoriana, né le posizioni appena esposte. Ciò che li caratterizza è il tentativo di costruire una cornice filosofica e metodologica in cui vi sia spazio per l’autonomia del livello di analisi dei fenomeni proprio della psicologia (da concepire come una condizione permanente e non come una situazione temporanea da accettare strumentalmente in attesa che si consolidi l’apparato concettuale destinato ad eliminarla) e, ad un tempo, per i nuovi strumenti di ricerca riassunti nel precedente paragrafo.

Per fare questo gli autori in questione partono da una riflessione epistemologica di

⁴ Anche le versioni del computazionalismo classico che non teorizzano l’esistenza di un linguaggio del pensiero, ma che comunque ritengono che una psicologia avanzata dovrà usare rappresentazioni definite al livello logico-informazionale, vengono considerate al più modelli teorici approssimativi che dovranno trasformarsi nel loro equivalente connessionista.

carattere generale il cui scopo è mostrare che cosa è possibile conservare oggi dell'ambizioso progetto di unificazione delle scienze sviluppato dalla tradizione neopositivista. Tale riflessione mira, come vedremo, a ripensare in modo radicale quel progetto affidandosi in primo luogo agli strumenti concettuali messi a disposizione dalla odierna filosofia della biologia.

Come è noto, il progetto neopositivista, fin dalle sue origini (si pensi al "manifesto" di Hahn, Neurath e Carnap del 1929) mirava a realizzare una ricostruzione unitaria della scienza attraverso lo strumento dell'analisi logica. Il modello nomologico-deduttivo (formalizzato successivamente) divenne essenziale in questo progetto, poiché mostrava come, in linea di principio, l'intero edificio del sapere scientifico potesse essere ricondotto alle strutture concettuali della fisica. Le forti critiche cui questo modello è stato sottoposto negli anni sessanta da Kuhn e Feyerabend partendo dalla tesi della incommensurabilità delle teorie scientifiche hanno poi portato a importanti revisioni (Schaffner 1967; Kitcher 1981).

Un elemento comune caratterizza comunque il modello nomologico-deduttivo e tutte le revisioni cui esso è stato successivamente sottoposto: *l'idea che la riduzione sia una relazione deduttiva tra teorie. E' esattamente questo presupposto che viene contestato dai teorici cui farò riferimento nelle prossime pagine* (William Wimsatt, Robert Richardson, William Bechtel, Carl Craver, Stuart Glennan, Peter Machamer, Adele Abrahamsen, Lindley Darden, per citarne alcuni).

William Wimsatt ha iniziato negli anni settanta (ma vedi anche Pattee 1970; Weiss 1971) a sviluppare una rilettura del concetto di riduzione che la concepisse *non come una operazione formale tra teorie, ma come una integrazione tra ambiti di ricerca operanti a diversi livelli di analisi dei fenomeni*, partendo dalla constatazione che

"at least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms, rather than explaining theories by deriving them from or deducing them to other theories, and this is seen by them as reduction, or as integrally tied to it" (Wimsatt 1972).

Lindley Darden e Nancy Maull (1977), su questa stessa linea, hanno messo a punto il concetto di "campo di ricerca" (traduco così l'espressione "field", introdotta inizialmente da Dudley Shapere), ritenendolo più adatto della nozione di "teoria" allo scopo di comprendere le reali dinamiche di sviluppo della ricerca scientifica in passato come oggi: il campo di ricerca viene definito nei seguenti termini:

“A central problem, a domain consisting of items taken to be facts related to that problem, general explanatory facts and goals providing expectations as to how the problem is to be solved, techniques and methods and sometimes, but not always, concepts laws and theories which are related to the problem and which attempt to realize the explanatory goals”
(Darden e Maull,1977)

Dietro questa posizione vi è una precisa idea su quale deve essere il ruolo della filosofia della scienza oggi: la sconfitta del progetto di individuare un criterio logico di demarcazione tra scienza e metafisica (sia esso di tipo verificazionista o falsificazionista) ha condotto alcuni autori (fra cui quelli che stiamo ora trattando) ad un tentativo di ricostruzione razionale dello sviluppo storico delle teorie scientifiche, sulla linea dei lavori di Lakatos e (forse ancora di più) di Laudan: il filosofo della scienza deve addentrarsi nelle articolazioni reali (non quelle auspiccate sulla base di un qualche criterio metastorico) della pratica scientifica, indagare sulla struttura concettuale che in essa si esprime, porre in evidenza e discutere le questioni metodologiche e filosofiche che tale struttura incorpora. La nozione di “campo di indagine” mira a svolgere esattamente questo ruolo. Un campo di indagine è definito da un problema centrale intorno al quale prendono forma modelli esplicativi, metodi di ricerca e anche teorie. Tra campi di indagine, secondo le autrici citate, possono esservi diversi tipi di relazione: esempi possono essere relazioni del tipo *struttura-funzione* (la chimica fisica indaga la struttura delle molecole mentre la biochimica indaga la funzione da esse svolta); del tipo *localizzazione fisica di entità funzionalmente definite* (i cromosomi identificati dai citologi nelle cellule costituiscono la localizzazione fisica dei geni teorizzati dai genetisti); del tipo *causa effetto* (le dinamiche biochimiche come causa dei pattern ereditabili di espressione genica). Lo sviluppo della scienza è spesso caratterizzato dalla integrazione tra diversi campi di ricerca, che sovente produce il costituirsi di “interfield theories”⁵.

Nei vari campi di ricerca un tipico oggetto di studio sono poi, come è emerso già dalla citazione di Wimsatt, i “sistemi meccanici” (“mechanisms”). Tale espressione non va ovviamente intesa nel senso del meccanicismo seicentesco: essa si riferisce a *qualsiasi sistema che espliciti una certa funzione in virtù delle sue parti componenti*,

⁵ (un esempio interessante del suddetto processo di integrazione che, negli anni trenta, collegò tra di loro i campi (prima nettamente separati) della ricerca sulle vitamine e della ricerca sulla “respirazione intracellulare”, portando alla identificazione delle vitamine come costituenti dei coenzimi si ha in Bechtel (1984).

delle loro operazioni e della loro organizzazione. L'organizzazione più semplice di un sistema è quella lineare, nella quale ogni parte di esso interviene sull'output di una parte che ha operato in precedenza, e così via lungo una catena lineare, finché non viene generato l'output globale del sistema. Nelle scienze della vita si lavora invece spesso con *sistemi non organizzati linearmente*, cioè sistemi in cui, per esempio, vi sono cicli nei quali l'attività di una certa componente, anziché causare l'attività di quella successiva in una certa sequenza, può retroagire su uno stadio precedente del processo. Un esempio fra i tanti possibili che ricorrono negli scritti degli autori citati è il ciclo di Krebs (il fondamentale ciclo metabolico proprio di tutte le cellule che utilizzano ossigeno nei processi di respirazione cellulare); ma biochimica, fisiologia ed ecologia hanno proposto una grande quantità di questi esempi, in cui entrano in gioco circuiti a retroazione positiva o negativa (vedi Craver 2001).

Bechtel sottolinea che ciò che fa un sistema meccanico (specie quelli non organizzati linearmente) nella sua globalità è ben diverso da ciò che fanno le sue componenti: le proprietà che nascono dalla organizzazione del sistema non sono in alcun modo deducibili da quelle delle sue parti. Per esempio, il comportamento di sequenze di amminoacidi considerate isolatamente rivela poco sul comportamento delle stesse sequenze quando sono "piegate" secondo le note modalità della struttura ternaria e quaternaria delle proteine. Tali comportamenti "globali" possono solo essere individuati empiricamente e non dedotti formalmente da qualsivoglia serie di leggi. Non bisogna poi dimenticare l'altro fattore essenziale nello studio dei sistemi meccanici, oltre alla loro organizzazione, cioè l'ambiente in cui essi operano: quei particolari sistemi meccanici che sono gli organismi biologici, infatti, si sono evoluti per operare in una certa tipologia di ambiente e quindi accade che alcune caratteristiche di quell'ambiente vengano "incorporate" nel funzionamento del sistema; esso, per esempio, può non aver bisogno di generare al proprio interno una certa sostanza perché la trova nell'ambiente circostante (si pensi al caso delle vitamine, disponibili per noi ed i nostri antenati in vari tipi di cibi; non vi è stata, ovviamente, alcuna pressione evolutiva perché emergesse una capacità da parte nostra o dei nostri antenati di sintetizzarle). In questo caso il funzionamento del sistema, necessitando esso del suddetto apporto esterno, potrà essere compreso in modo soddisfacente indagando, oltre che la sua organizzazione interna, il contesto ambientale circostante.

Una analisi di questo tipo conduce Bechtel (ma anche Craver 2001; Glennan 1996) a sottolineare che *l'inconsistenza del modello nomologico-deduttivo non sta solo nel*

fornire una inadeguata ricostruzione della pratica scientifica, ma nel non comprendere la struttura della spiegazione scientifica, che solo attraverso una indagine puntuale e rigorosa di quella pratica può emergere. La distinzione tra livelli di analisi dei fenomeni è una caratteristica strutturale della ricerca scientifica e non una condizione provvisoria da accettare in attesa che si realizzi l'evento delle riduzione di tutti i fenomeni ad un livello fondamentale.

Bechtel e Hamilton (2007) mostrano l'approssimazione che inevitabilmente caratterizza il modello nomologico-deduttivo quando viene applicato alle scienze biologiche: il cosiddetto "dogma centrale" della biologia, cioè l'idea che la sequenza lineare DNA-RNA-proteine possa essere la chiave di volta per comprendere tutti i fenomeni biologici si è da tempo rivelato un mito. Oggi sappiamo che il DNA, senza il resto di quel complesso sistema che è la cellula (proteine e, in particolare, enzimi) non è in grado di fare nulla; Richard Lewontin (2000) ha detto, con una immagine molto efficace, che, oltre al DNA che *specifica* il codice necessario per le varie sostanze che servono all'organismo, occorre qualcosa che *legga* quel codice, qualcosa che *costruisca* sulla base di ciò che è stato letto e qualcosa che *scriva* di nuovo il codice per altre iterazioni del ciclo; nessuna delle ultime tre operazioni può essere svolta dal DNA. Sappiamo poi che il complesso fenomeno della regolazione dei geni richiede molte informazioni relative all'ambiente (il sistema cellulare) in cui il gene opera, col che risulta evidente che occorre appoggiarsi alle strutture concettuali della chimica-fisica e non solo della genetica.

Questa linea di pensiero, pur teorizzando l'autonomia dei vari livelli coinvolti nella spiegazione del comportamento di un sistema meccanico, *non condivide affatto, tuttavia, la tesi della realizzabilità multipla del mentale* (uno stato di dolore può essere proprio di un uomo, di un polipo e di un ipotetico sistema artificiale che condividano una stessa descrizione funzionale; mentre, evidentemente, i relativi substrati fisici sono assai eterogenei tra di loro): essa è infatti incompatibile con la maggior parte delle strategie di ricerca che hanno prodotto successi nelle scienze biologiche. I biologi, pur sapendo che i sistemi meccanici che veicolano certe funzioni possono variare ampiamente nelle diverse specie, utilizzano spesso le conoscenze possedute relativamente ad una specie per incrementare le conoscenze relative ad un'altra specie: basti pensare alla estrema utilità che ha avuto lo studio della corteccia visiva dei macachi per comprendere il funzionamento del sistema visivo negli organismi umani. La convinzione che ha ispirato simili strategie di successo è che, pur nel quadro delle suddette differenze interspecifiche, esistano importanti aspetti comuni

relativi ai vari sistemi meccanici nelle diverse specie: questo perché i sistemi biologici tendono, dove possibile, a conservare alcune caratteristiche strutturali nel corso della evoluzione; i mutamenti più radicali hanno infatti costi particolarmente elevati in termini di capacità adattative.

Ma c'è un'altra riflessione che rende la tesi della realizzabilità multipla poco plausibile: William Bechtel la espone con molta chiarezza:

Those who view multiple realizability (qui il riferimento è a i rapporti tra neuroscienze e psicologia, ndr) as an obstacle for reduction, often neglect a further factor: just as there are neural differences between organisms and especially between species, there are psychological differences as well. The behavior of a hungry octopus is very different from that of a hungry human. (Bechtel 2007)

Detto in altri termini: nello studiare un certo processo sia lo psicologo che il neuroscienziato possono scegliere una analisi "a bassa risoluzione" (coarse-graned) del sistema meccanico indagato oppure una analisi "ad alta risoluzione" (fine-graned): ciò in funzione degli specifici obiettivi che la loro ricerca si pone in quel momento. E' dunque possibile adottare una definizione funzionale di "fame" a bassa risoluzione, che consente di applicare lo stesso predicato psicologico a diverse specie viventi (uomo e polpo, per restare all'esempio proposto); a quel punto, però, dovrà essere a bassa risoluzione anche l'analisi dei sistemi neurali presenti nelle due diverse specie, cioè sarà opportuno porre in evidenza le analogie strutturali esistenti tra di esse senza entrare nel dettaglio delle indubbe differenze che vi sono. E così sarà opportuno fare scegliendo l'opzione ad alta risoluzione: andranno poste in evidenza non solo le notevoli differenze nel substrato neurologico dei processi rappresentativi relativi alle due specie, ma anche le specifiche articolazioni che quei processi assumono alla luce di una dettagliata analisi del comportamento degli organismi in questione. *Gli esempi di realizzabilità multipla sembrano invece derivare dal fatto che si contrappone ad una analisi ad alta risoluzione del substrato neurologico una analisi a bassa risoluzione delle funzioni psicologiche.* Di norma, quanto più dettagliata diventa l'analisi ad entrambi i livelli quanto più ci si rende conto della scarsa plausibilità della tesi della realizzabilità multipla.

Il rapporto tra psicologia e neuroscienze, in questo quadro, viene drasticamente ripensato, sia rispetto al modello del "computazionalismo classico" che rispetto alle più radicali affermazioni della scienza cognitiva post-classica. La psicologia cognitiva

opera sul piano di analisi delle proprietà globali dei sistemi neurali, proprietà che non sono deducibili dalle caratteristiche dei singoli elementi di quei sistemi. Detto in altri termini, i sistemi meccanici presenti nel cervello *possono* produrre attività che risultano ben comprensibili nelle loro caratteristiche generali solo al livello funzionale-simbolico; ma, per spiegare *come* esse producono quelle attività diventa necessario utilizzare un modello connessionista. Il livello di descrizione simbolico non viene ritenuto per definizione implausibile, come certi autori precedentemente citati sostengono, poiché è talvolta in grado di descrivere le prestazioni di un organismo in una misura che può essere più soddisfacente rispetto ad un modello connessionista. D'altra parte esso non deve essere ipostatizzato come se si trattasse dell'unico livello al quale risulta intellegibile il comportamento umano. Per certi fenomeni può risultare più soddisfacente l'approccio fondato sulle reti neurali. E' insomma possibile riconoscere l'utilità di certi modelli simbolici senza dover adottare l'ipotesi generale di un inconscio cognitivo come unico paradigma di studio del mentale.

Bechtel mostra, in particolare, come lo studio di alcuni fenomeni relativi all'apprendimento linguistico, quali il consolidarsi del meccanismo che presiede alla formazione del tempo passato, mostri la *proficuità di una competizione tra modelli simbolici e connessionisti*, poiché entrambe le scuole di pensiero sono state indotte a produrre una caratterizzazione sempre più dettagliata del fenomeno in oggetto e previsioni sempre più precise da sottoporre a test sperimentali.

Oggi, specialmente nel settore degli studi sul linguaggio, capita di incontrare lavori comuni di scienziati "connessionisti" e "simbolisti" che, abbandonando le iniziali contrapposizioni, cercano di combinare modelli fondati su quasi-regole (rule like constraints) e modelli connessionisti (si vedano gli studi sulla capacità di sillabare nelle varie lingue in Prince and Smolensky 2004).

Questi ultimi rilievi, che mostrano come nel panorama della odierna scienza cognitiva non vi siano solo radicali contrapposizioni, ma spesso anche soluzioni metodologiche intermedie, *acquistano un senso particolare alla luce della prospettiva che abbiamo delineato*: l'analisi "simbolica" e quella connessionista non si escludono reciprocamente, perché la seconda abbisogna di una scomposizione delle varie funzioni cognitive che (in molti casi) solo la prima può dare; d'altra parte alla prima giova sicuramente poter fondare i propri modelli funzionali su una simulazione che mostri come i vari moduli via via delineati siano esprimibili da un sistema dotato di un minimo di plausibilità dal punto di vista neurobiologico.

Ciò che è auspicabile nello studio di ogni sistema meccanico è una *coevoluzione tra*

le due teorie che lo studiano al livello delle singole parti (le singole aree cerebrali coinvolte in un certo compito, modellizzate da una qualche rete neurale) ed al livello delle sue dinamiche funzionali complessive (riconducibili alla organizzazione delle parti e al rapporto del sistema con l'ambiente circostante). Questa *dialettica tra scomposizione funzionale e localizzazione* si è spesso rivelata positiva nel quadro delle odierne ricerche sulla mente-cervello: basti pensare, per esempio alle più recenti tecniche di localizzazione funzionale usate in neurobiologia, che presuppongono la scomposizione funzionale dei compiti studiati, ottenuta con il "tradizionale" metodo della psicologia cognitiva. Oppure agli sviluppi prodotti dalla teoria neurobiologica relativa alla organizzazione delle aree visive extrastriate, che ha influenzato in modo significativo i modelli psicologici della visione (vedi Marraffa 2002)

Lo studio dei vari sistemi meccanici in cui si articola l'attività mentale umana necessita dunque sia del livello della analisi funzionale che di quello che descrive la struttura, perché proprio in questo modo si possono realizzare feconde forme di coevoluzione fra i due piani, e quindi quei riorientamenti concettuali che portano talvolta a "interfield theories". La posizione fodoriana è giudicata erronea in quanto ipostatizza il livello di analisi simbolica ritenendolo l'unico in grado di produrre una spiegazione intellegibile dei nostri comportamenti. D'altra parte, i più radicali sostenitori della svolta connessionistica hanno compiuto l'errore di trasformare la loro legittima critica alla suddetta ipostatizzazione nella discutibile critica di ogni descrizione funzionale (dei processi mentali) che non sia riconducibile all'operato di una rete neurale.^{6 7}

⁶ A questo proposito vale la pena sottolineare che le scienze cognitive, così come le abbiamo presentate in questa terza parte, non sono in antitesi con la prospettiva ecologica in psicologia difesa da J.J.Gibson e dal secondo Neisser (1976): Gibson ha studiato la percezione umana come processo in cui *l'intero organismo umano si relaziona direttamente all'ambiente* in cui è immerso, individuando le varie tipologie di informazione cui il sistema visivo è sensibile; nel condurre una analisi a questo livello, non ha avuto alcun bisogno di concentrarsi sui sottosistemi che rendono possibile quella relazione ecologica. Su questa base ha anche criticato la psicologia da laboratorio, in cui vengono talvolta proposti stimoli troppo semplificati rispetto alle reali condizioni di vita dell'organismo umano, che si confronta con stimoli complessi, caratterizzati da estensione spaziale e durata temporale. Bechtel condivide la posizione di Gibson fin dove essa sottolinea l'importanza di non perdere di vista il quadro biologico ed evolutivo in cui l'organismo umano si colloca. In questo senso una moderna teoria della mente dovrà comprendere anche le osservazioni che abbiamo effettuato nella seconda parte di questo lavoro, quando abbiamo parlato di una cognizione estesa sul mondo. Ma questo rilievo non destituisce di valore lo studio a vari livelli delle diverse modalità rappresentative (non solo simboliche) con cui la mente umana si relaziona al mondo; l'antitesi tra l'antirappresentazionalismo di Gibson (e dei sostenitori della cosiddetta mente estesa) e la scienza cognitiva può essere interpretata nei termini di un inutile contrapporsi di due distinti livelli di analisi dei processi mentali (una tesi simile a quella di Bechtel. in Clark 1997).

⁷ L'approccio delineato in questa parte implica anche una precisa risposta alle teorie che, partendo dalla constatazione del carattere situato della cognizione umana, concludono che la mente umana va "estesa al mondo" perché qui sono le impalcature materiali senza le quali i nostri processi cognitivi non potrebbero avere luogo. Bechtel si rifa ad alcune recenti riflessioni nel campo della biologia teorica (Ruiz-Mirazo, Pereto e Moreno 2004),

Un'ultima passo va fatto in relazione alla più generale cornice filosofica che questa proposta metodologica implica: la teoria della mente fodoriana contiene una precisa (anche se discutibile) trattazione del problema dell'intenzionalità. La scienza cognitiva post-classica tende ad una posizione eliminativista. La prospettiva delineata in quest'ultima parte, che potremmo denominare di *naturalismo biologico*, implica una proposta di soluzione radicalmente diversa. Riassumo qui, senza argomentare nel dettaglio, le riflessioni di Bechtel (1988):

1) viene proposta una definizione *esternalistica* dell'intenzionalità: gli stati intenzionali sono stati relazionali che si costituiscono tra un organismo (non necessariamente solo umano) ed il suo ambiente: per un organismo avere una credenza sull'acqua significa stare in una appropriata relazione con l'acqua (qui si nota l'influenza della teoria del significato di Putnam); l'ambiente entra dunque a far parte delle caratteristiche definitorie dei processi mentali.

2) Il problema, presente soprattutto in un approccio esternalistico, relativo al fatto che gli stati intenzionali possono rappresentare oggetti non esistenti, viene risolto con il concetto di "mondo nozionale" ("notional world"), un concetto presente nelle riflessioni di Dennett (1987), che Bechtel utilizza però a suo modo: si tratta di ogni mondo possibile definito a partire da quello attuale considerando vere le credenze false di un certo organismo e ragionevoli i suoi desideri non attuabili. Gli stati mentali di un organismo sono comparabili ai tratti biologici: come alcuni di essi sono ben adattati ed altri no (non tutte le parti che ci costituiscono come organismi esistono in quanto selezionate dall'ambiente; in questo caso possiamo dire in quale tipo di ambiente queste parti sarebbero adatte), così alcune credenze saranno vere, cioè i relativi stati mentali avranno una precisa relazione col mondo; mentre altre saranno false, avendo un riferimento solo in uno dei suddetti "notional worlds". *La relazionalità col mondo che abbiamo assunto come caratteristica definitoria degli stati intenzionali va dunque intesa in senso olistico*, riferendola all'intero insieme degli stati cognitivi di un organismo, perché è evidente che le singole credenze false non possono avere da sole quella proprietà relazionale (non si può stare in una relazione con qualcosa

che considerano ogni organismo vivente come un *sistema autonomo* riprendendo una definizione di Francisco Varela e Humberto Maturana: ogni organismo vivente è un sistema lontano dall'equilibrio termodinamico e continuamente soggetto a pressioni ambientali che tendono a dissipare verso l'esterno l'energia che esso ha immagazzinato a vari livelli ed in varie forme. Per questo gli organismi devono sviluppare *forme di segregazione dei propri processi* rispetto all'ambiente e queste forme di segregazione si fanno via via più complesse quanto più si sale nella scala evolutiva, determinando confini parzialmente permeabili ma ben definiti tra un sistema meccanico e l'altro. L'idea della mente estesa, dunque, se interpretata in maniera forte, sembra andare contro una caratteristica strutturale degli esseri viventi.

che non esiste).

3) La conoscenza della relazione in cui l'organismo sta col suo ambiente non va confusa con quella relativa alle specifiche modalità di funzionamento dei processi che rendono possibile quella relazione: in questo senso il livello di descrizione "intenzionale" non può essere confuso con i vari livelli di elaborazione dell'informazione (simbolici, connessionistici). L'organismo vivente, in quanto sistema meccanico, ha delle proprietà globali (la relazione col mondo in questione) che non possono essere comprese concentrandosi sulle sue varie parti (appunto i vari sottosistemi meccanici che lo costituiscono). Gli stati intenzionali ci dicono quali aspetti del sistema-organismo necessitano di una spiegazione: i tipi di informazione cui un organismo è sensibile, i tipi di cose che può ricordare, i modi in cui può usare le precedenti informazioni.

Propositional attitudes are a way of characterizing the cognitive system vis a vis its environment, but it is not uncommon in science to use different accounts to describe the behavior of a system and to describe the internal processes that make the behavior possible. For example, a yeast cell performing fermentation is described physiologically as metabolizing sugar to produce alcohol, whereas in biochemistry the reaction is explained in terms of networks of enzymes and cofactors which together make it possible for a cell to metabolize the sugar. Similarly, we can view the characterization of how a system relates to its environment as different from the processing model that explains how it is able to accomplish this. (Bechtel 1988b)

4) L'universo degli atteggiamenti proposizionali (credenze, desideri etc.) definisce dunque un insieme assolutamente essenziale, seppur molto generale, di dati sulla relazione dell'organismo col mondo (naturale e sociale). Siamo di fronte ad una prospettiva di realismo intenzionale, che rifiuta qualsiasi lettura strumentalistica degli atteggiamenti proposizionali e, a fortiori, il materialismo eliminativo.

CONCLUSIONI

L'approccio che ho tratteggiato mira:

1) ad una riflessione di ordine epistemologico che riprende aspetti del lavoro di Lakatos, Laudan e comunque di quei filosofi della scienza rivolti a compiere una ricostruzione razionale della storia della scienza capace di farla interagire proficuamente con la pratica scientifica odierna.

2) Questa operazione si concentra soprattutto sulle scienze biologiche, ritenendo che i problemi del progetto di unificazione delle scienze di marca neopositivistica nascessero soprattutto da un eccessivo “schiacciamento” sullo studio della fisica.

3) Proprio questa scelta porta a proporre un approccio metodologico che, più che unificare la scienza, suggerisce delle strade attraverso cui ottenere forme di integrazione tra conoscenze collocabili a diversi livelli. Ciò nel quadro di una prospettiva in cui, più che teorie, ci sono campi di ricerca e, più che leggi formalmente definite, ci sono sistemi meccanici da studiare in una ottica di coevoluzione tra modelli teorici distinti, che può portare anche alla nascita di “interfield theories”. In questa cornice teorica prende forma una teoria della mente che mira a fornire una prospettiva unitaria al convulso proliferare di modelli nelle scienze cognitive.

Come ogni prospettiva unificatrice, essa rischia in qualche punto di apparire un po' generica; certamente necessita di ulteriori articolazioni. Ma mi sembra, nonostante tutto, la proposta più convincente tra il solipsismo metodico di Fodor ed il materialismo eliminativo di molti entusiasti sostenitori della scienza cognitiva post classica.

BIBLIOGRAFIA

ABRAHAMSEN A. and BECHTEL W. (2006). Phenomena and mechanisms: putting the symbolic, connectionist and dynamical system perspective, in R. Stainton (ed), *Contemporary debates in cognitive science*, Oxford, Basil Blackwell.

BECHTEL W. (1984). Reconceptualizations and interfield connections: the rediscovery of the link between vitamins and coenzymes, *Philosophy of science* , 51, 265-292

BECHTEL W. (1988b). *Philosophy of mind*, Hillsdale, NJ, Erlbaum

BECHTEL W., HAMILTON A. (2007). Reduction, integration and the unity of sciences. In: T. Kuipers (ed), *Philosophy of science: focal issues* (vol 1 of the Handbook of the Philosophy of Science), New York, Elsevier.

BECHTEL W. (2007). Reducing psychology while maintaining its autonomy via mechanistic explanation. In: M. Schouten e H. Looren de Jong (eds). *The matter of mind: philosophical essays on psychology, neuroscience and reduction*, Oxford, Basil Blackwell.

BECHTEL W. (2008). Explanation: mechanism, modularity and situated cognition. In: P. Robbins and M. Aydede (eds). *Cambridge handbook of situated cognition*, Cambridge, Cambridge University Press

BROOKS R. (1991). Intelligence without representation, *Artificial Intelligence*, 47, 139-159.

CASALEGNO C. (1998). *Filosofia del linguaggio*, Roma, Carocci.

CHURCHLAND P. (1995). *The engine of reason, the seat of the soul*, Cambridge, M.I.T Press

CHURCHLAND P.S., SEJNOWSKI T.J. (1992). *The computational brain*, Cambridge, MIT Press.

CLARK A. (1997). *Being there*, Cambridge, MIT Press

CRAVER C. (2001). Structures of scientific theories. In: P.K. Machamer and M. Silberstein (eds): *Blackwell guide to the philosophy of science*, Oxford, Blackwell

DARDEN L. and MAULL N. (1977). Interfield theories, *Philosophy of science*, 43, 44-64.

DENNETT D.C. (1987). *The intentional stance*, Cambridge, MIT Press

DREYFUS H.L. (1972). *What computers can't do*, Harpers and Row Publishers Inc, New York.

FODOR J.A. (1975). *The language of thought*, New York, Thomas Y. Cromwell

FODOR J.A. (1980). Methodological solipsism considered as a research strategy in cognitive psychology, *Behavioral and brain sciences*, 3, 69-109.

FODOR J.A. (1987). *Psychosemantics*, Cambridge, MIT Press

GIBSON J.J. (1979). *The ecological approach to perception*, Boston, Houghton Mifflin.

GLENNAN S. (1996). Mechanisms and the nature of causation, *Erkenntnis*, 44, 50-71.

HAHN H, NEURATH O, CARNAP R.(1929). *Wissenschaftliche Weltauffassung der Wiener Kreis*. Trad. Italiana: *La concezione scientifica del mondo*, Roma-Bari 1979.

HULL D. (1974). *Philosophy of biological science*, Prantice Hall, Englewood Cliffs.

KITCHER P. (1981). Explanatory unification, *Philosophy of science*, 48, 507-531.

LAUDAN I. (1987). Progress or rationality? The prospects for normative naturalism, *American Philosophical Quarterly*, 24, 19-31.

MARRAFFA M. (2002). *Scienza Cognitiva: una introduzione filosofica*, Padova, CLEUP

MACHAMER P., DARDEN L., CRAVER C. (2000). Thinking about mechanisms, *Philosophy of science*, 67, 1-25.

NAGEL E. (1960). *The structure of science*, New York, Harcourt

NEISSER U. (1976). *Cognition and reality*, San Francisco, Freeman

PATTEE H.H. (1970). The problem of biological hierarchy. In: Waddington C.H. (ed.): *Towards a theoretical biology*, Edinburgh, 117-135.

PIATTELLI PALMARINI M.(2008). *Le scienze cognitive classiche: un panorama*, Torino, Einaudi 2008

PRINCE A., SMOLENSKY P. (2004). *Optimality theory: constraint interaction in generative grammar*, Oxford, Backwell

RUMELHART D., McCLELLAND J. et al. (1986). *Parallel distributed processing*, Cambridge, MIT Press.

RICHARDSON R.C. (1982). How not to reduce a functional psychology, *Philosophy of science*, 49, 125-137.

RUIZ-MIRAZO K., PERETO J., MORENO A. (2004). A universal definition of life: autonomy and open-ended evolution. In: *Origins of life and evolution of the biosphere*, 34, 323-346.

SCHAFFNER K.F. (1967). Approaches to reduction, *Philosophy of science*, 34, 137-147.

SHAPER D. (1974). Scientific theories and their domains. In: F. Seppé (ed), *The structure of scientific theories*, Urbana, University of Illinois Press.

WEISS P.A. (1971). *Hierarchically organized systems in theory and praxis*, New York, 1971.

WIMSATT W.C. (1976). Reductive explanation: a functional account. In: Cohen R.S. et al (eds), *PSA 1974*, Dordrecht, 671-710.